

AD-A246 945



Average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

ATE

3. REPORT TYPE AND DATES COVERED

ANNUAL 15 Apr 90 TO 14 May 91

4. TITLE AND SUBTITLE

PSYCHOLPHYSICAL ANALYSES OF PERCEPTUAL REPRESENTATIONS

5. FUNDING NUMBERS

AFOSR-90-0274

PE - 61103D

PR - 3484

TA - HS

6. AUTHOR(S)

Dr Irving Biederman

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Minnesota
Department of Psychology
75 East River Road
Minneapolis, MN 55455

DTIC
ELECTE
MAR 05 1992
S D

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Dr Tangney
AFOSR/NL
Building 410
Bolling AFB DC 20332-6448

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

AFOSR-TL- 92 0105

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release;
distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

A number of collaborative projects have been launched during the first year and a half of the grant. Consistent with the original proposal, the common theme to all these projects is the linkages between early sensory (psychophysical) processes and perceptual representations that provide access to cognition. The individual projects are summarized in the body of the report. In addition to the research projects, two informal weekly seminars were held throughout the first year among Center personnel and those with closely related interests. One was concerned with the development of a front end for the kind of object recognition model described by RBC. The other was a general examination of recent research in neural net type models.

This document has been approved
for public release and sale; its
distribution is unlimited.

14. SUBJECT TERMS

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT

(U)

18. SECURITY CLASSIFICATION OF THIS PAGE

(U)

19. SECURITY CLASSIFICATION OF ABSTRACT

(U)

20. LIMITATION OF ABSTRACT

(U)

PROGRESS REPORT

Submitted to:

Air Force Office of Scientific Research
AFOSR/NL
Building 410
Bolling A.F.B., D.C. 20332-6448

Dr. John Tangney, Program Manager

Title:

PSYCHOPHYSICAL ANALYSES OF PERCEPTUAL
REPRESENTATIONS

Grant No.

AFOSR-90-0274

Date

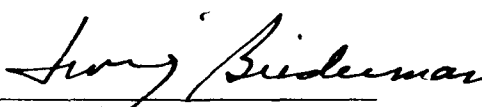
3 DEC 1991

Submitting
Organization:

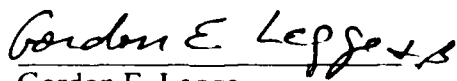
University of Minnesota
Office of Research Administration
1919 University Ave.
St. Paul, MN 55104

PIs:

Professor Irving Biederman
(213) 740-6094


Irving Biederman
134-28-9789

Professor Gordon E. Legge
(612) 625-0846


Gordon E. Legge
023-42-1595

Department of Psychology
University of Minnesota
Elliott Hall
75 East River Road
Minneapolis, MN 55455

Accession For	
NTIS CRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

92-05707


92 3 03 184

**Progress Report for AFOSR Grant No. AFOSR-90-0274
Psychophysical Analyses of Perceptual Representations**

A number of collaborative projects have been launched during the first year and a half of the grant. Consistent with the original proposal, the common theme to all these projects is the linkages between early sensory (psychophysical) processes and perceptual representations that provide access to cognition. The individual projects are summarized in the body of the report. In addition to the research projects, two informal weekly seminars were held throughout the first year among Center personnel and those with closely related interests. One was concerned with the development of a front end for the kind of object recognition model described by RBC. The other was a general examination of recent research in neural net type models.

Personnel supported during the 1990-91 academic year:

We were successful in recruiting three outstanding post doctoral fellows during the first year of the grant: David Knill (Ph. D. Brown University), J. Stephen Mansfield (Ph. D. Oxford University), and John Hummel (Ph. D. University of Minnesota). Knill and Mansfield are continuing for the 1991-1992 academic year; Hummel has accepted a position as an Assistant Professor of Psychology at UCLA. In addition, two outstanding graduate students were recruited for the grant: Pascal Mamassian (undergraduate degree, Paris) and Zli Liu (Undergraduate degree, University of Peking).

RESEARCH PROJECTS

Structure-from-motion based on surface boundaries (Thompson, Kersten, & Knecht)

Existing computational models of structure-from-motion -- the appearance of three-dimensional motion generated by moving two-dimensional patterns -- are all based on variations of optical flow or feature point correspondences within the interior of single objects. Bill Thompson, Dan Kersten, and a graduate student Bill Knecht investigated a pattern in which a rotating cylinder is seen, though no variation in optical flow exists across the apparent cylinder. They argued that the perception of a cylinder is based on cues present at the discontinuities of flow seen as boundaries of the cylinder. A simple computational model sufficient to describe this effect was proposed. The model is based on qualitative constraints relating possible object motions to patterns of flow, together with an understanding of the patterns of flow that can be discriminated in practice.

Thompson, W. B., Kersten, D., & Knecht, W. R. (1991 submitted). Structure-from-motion based on information at surface boundaries. Submitted to Biological Cybernetics.

Shadow Contours (Mamassian, Knill, & Kersten)

Shadow contours provide a potentially rich source of information for surface shape. Though a number of studies have been done on the perceptual interpretation of shadow contours, they have typically proceeded without any formal analysis of the structure of shadow boundaries. Graduate student, Pascal Mamassian, post-doc Dave Knill, and Dan Kersten derived a number of constraints on the relationship between shadows and their causes.

A natural way to categorize different types of shadow boundaries is according to the means by which they were formed. Attached shadow boundaries are formed at points of a surface at which the surface curves away from the lighting direction. Cast shadow boundaries demarcate regions of a surface which are occluded from the light source by a distal surface patch. The

different physical processes which underlie the formation of the two types of shadow boundary lead to different sets of constraints on their projected contours. Furthermore, the shadow formation processes lead to two qualitatively different kinds of constraints: those dealing with the geometrical structure of the contours and those dealing with the luminance pattern in the immediate neighborhood of the contours. These constraints, as well as constraints on the behavior of shadow contours at intersections with other types of contour in an image (e.g. smooth self-occlusions and surface creases) were analyzed.

However attractive the formally derived properties may be, some of them may not be of practical use to the human visual system. Therefore a number of "soft" constraints were proposed, which are only guaranteed to hold most of the time. Psychophysical investigations are now beginning to address which constraints are in fact incorporated into visual system processing. In one experiment, they are investigating whether the convexity or concavity of a surface patch is influenced by the cotangency of the occluding contour with the attached shadow.

Some of this research has been accepted for presentation:

Mamassian, P., Knill, D. C., & Kersten, D. (1991 accepted). What's behind the shadow contours? Association for Research in Vision and Ophthalmology, Sarasota, Florida.

Moving Cast Shadows (Kersten, Mamassian, & Knill)

The distance between the image of an object and its cast shadow provides information about the distance between the object and the shadowed surface. In principle, this information could be used to build up, over space or time, a perceptual model of the spatial relationship of the object casting the shadow to the underlying surface. Kersten, Mamassian, and Knill have constructed static and dynamic demonstrations in which the convexity or concavity of a shape theoretically could be disambiguated by the distance between the image of an object and its cast shadow. This will only work if the object boundary is assumed to be straight, in the static case, or is moving on a linear trajectory in the dynamic case. If the perceived shape is not affected, the remaining consistent interpretation would be to see straight lines and trajectories as curved.

In the static displays, a straight edge is made to cast a curved shadow on an ambiguous surface that can appear concave or convex. The perceived shape is not easily affected by the form of a static cast shadow boundary relative to the straight edge. The straightness of the edge of the casting object still appears straight. Incorrect shadow geometry is tolerated by the visual system over a wide range, and global consistency is very weakly enforced, if at all.

When a ball moves in a linear trajectory across a surface, the image trajectory of the cast shadow is determined by the shape and orientation of the underlying surface. A demonstration shows that the perceived path of the ball is strongly affected by the relationship of its image to its cast shadow. Linear trajectories can appear curved, non-bouncing balls appear to bounce, and stationary objects appear to move. Although the perceived path of the ball is affected by its cast shadow, the shadow information does not necessarily disambiguate the perceived shape of the underlying surface.

This work has been accepted for presentation:

Kersten, Daniel, Mamassian Pascal, and Knill, David (1991 accepted) Moving Cast Shadows Generate Illusory Object Trajectories. Association for Research in Vision and Ophthalmology, Sarasota, Florida.

Statistical efficiency for 3D recognition of wire objects. (Kersten, Knill, & Liu)

Models of visual recognition are often distinguished by being primarily image or object based. Alternatively, models of recognition can be classified according to the amount of "3D image understanding" incorporated in the model--that is, the degree to which three dimensional constraints are used in recognition.

With Knill and Kersten, graduate student Liu applied a novel paradigm to investigate the degree to which human observers incorporate 3D constraints in object recognition. They derived two ideal observers for a number of different recognition tasks. The first--the 3D ideal--knows that the image came from a projection of an object that may have been rotated and translated. The performance of any model must be less than or equal to this ideal. The second, 2D ideal, lacks 3D understanding in that it doesn't know that new views could be derived from 3D rotations of the same object, but does know that the images could have been rotated, translated or reflected in 2D. Observers that incorporate 3D constraints may actually beat this ideal at a recognition task. The signal domain consisted of views of 3D wire objects. The observation domain consisted of images of the same wire objects with vertices perturbed by 3D Gaussian noise.

Human efficiencies were calculated relative to both ideal observers for object detection and recognition tasks. The experiments show that human statistical efficiencies relative to the 2D ideal can exceed 100%. Any efficiencies over 100% indicate processes that either implicitly or explicitly make use of 3D understanding. The efficiencies relative to the two ideals can be used to quantify the degree to which the 3D constraints are incorporated in the recognition process.

This work was presented at ARVO:

Liu, Z., Knill, David C., Kersten, D. (1991) Statistical efficiency for 3D recognition of wire objects. Association for Research in Vision and Ophthalmology, Sarasota, Florida.

Efficiency of Judgments of Nonaccidental vs. Metric Differences in Contour (Mansfield, Biederman, Legge, & Knill)

Mansfield, Biederman, Legge, and Knill have been working on a project investigating the processing of non-accidental properties in the image (as described in Recognition By Components) in human vision. The RBC model of object recognition suggests the importance of certain contour shape primitives, such as straightness, in the recognition process. In a sense, these shape primitives form an alphabet from which visual representations of objects are formed. For an initial study it was chosen to measure the efficiency with which lines of different curvatures are processed - the prediction being that lines with the non-accidental property of straightness would be preferentially processed over lines that were curved.

Currently, this prediction is being tested using psychophysical procedures. The aim is to measure the statistical efficiency with which observers can categorize curves. In the experiment, the observer is made familiar with (trained upon) two categories of curved lines. The categories are described by Gaussian functions, each defined by a mean curvature and a standard deviation. Typically these parameters are chosen so that there is a considerable overlap between the two categories. In subsequent experimental trials the observer is shown a series of curves randomly selected from the two experimental categories. The observer is required to indicate the category from which each curve is selected. The accuracy with which the observer can perform the categorization task can be compared with the performance of an ideal detector to yield a measure of efficiency.

Data have been collected for two conditions. One where the orientation of the tangent at the mid-point of each curve is fixed, and one where the orientation of each curve is free to vary. With fixed orientation the sign of curvature (either positive or negative) was clear to the observer and it

was possible to use categories from either side of zero-curvature (a straight line). With randomized orientation, the sign of curvature cannot be determined, and the observer is only able to respond to absolute curvature. For the former condition a large data set has been collected from one observer (JSM), and partial data sets have been recorded from two naive observers. Measurements have been made with categories including zero-curvature (a straight line), categories including lines which curve through 180 degrees (U curves), and with 7 other intermediate category sets. Preliminary analysis of these data indicates that the observer is more efficient (efficiency = ~65%) at categorizing curves when the two categories were either side of zero curvature (i.e., where the subject's criterion would ideally be a straight line). Efficiencies were lowest for the condition where the ideal criterion was a line that curved through just less than 90 degrees. In the other data sets these trends are not so clear. One problem with this experimental approach is that the data are inherently noisy, and large data sets are required to produce repeatable results. Further data collection is in progress.

Two further subjects have been run with trials where the orientation of the curve was randomized. For these subjects the data show clear trends. As in the fixed orientation case, efficiencies are greatest for curvatures around zero. Efficiencies decrease down to a minimum for lines that curve through just less than 90 degrees. For even greater curvatures efficiencies improve (although not up to the level of near zero curvatures). Also, as the mean curvature increases, the preliminary analysis also shows an increase in mean reaction times for making the categorization. It is possible that the increase in efficiency as curvatures increase beyond 90 degrees represents the serendipitous use of another nonaccidental contrast -- parallelism of the two halves of the contour.

A confounding factor in these data is that observer's discriminability of curvature is better near zero curvature and decreases for greater curvatures. This is a common result which has been confirmed in curvature discrimination experiments using the stimuli and apparatus of the above study. To be able to conclude that the changes in efficiency are a result of the subjects' ability to maintain a stable internal criterion, it is necessary to account for variations in efficiency that might be produced by the sensory apparatus discriminating the curves. Methods by which these effects could be accounted for include determining the variability added to the stimulus by the curvature discriminatory processes (by measuring curvature discrimination), or by increasing the variability in the experimental stimuli to be much greater than that which might be added by curvature discrimination.

Together with completing the current data collection and analysis, this project is currently investigating the effects of altering the spread of the experimental categories. Code is also being written to perform similar experiments directly investigating the efficiency for recognizing lines that are parallel.

This work was presented at the Meetings of the Optical Society in November.

Texture Segregation of Patterns Distinguished by Conjunctions of Orientation and Spatial Frequency (Biederman, Legge, & Luebker).

There has been widespread interest in the failure, documented by Treisman, of humans to effortlessly search for targets defined by conjunctions (e.g., a red N among blue Ns and red Os) or to effortlessly perceive the border between texture fields defined by such conjunctions (e.g., between a region composed of a mixture of red Ns and blue Os and a region composed of red Os and blue Ns). Search RTs for conjunctive targets increase linearly with the number of distractors. In both the search and texture task, it is clear that a target or texture region defined by only a single feature, say a red letter among blue distractors, readily "pops out," producing what Julesz terms "effortless perception," in that there is virtually no increase in RTs as a function of the number of distractors. Conjunctive texture borders require much more time for their detection than borders defined by only a single feature.

Treisman has argued that the conjoining of independent attributes requires an act of attention that must be applied serially to each position in the visual field. However, there is ample physiological and psychophysical evidence that some conjunctions are performed automatically and in parallel in the visual field: Cells in V1 are tuned to particular combinations of orientation and spatial frequency. Would texture regions defined by conjunctions of orientation and spatial frequency be effortlessly combined? A negative answer to this question was provided by an experiment by Walters, Biederman, and Weisstein (1983). They showed that RTs for the detection of a border (as vertical or horizontal) in a 5 X 5 array of Gabor patches were greatly increased when the border was defined as a conjunction of spatial frequency and orientation compared to when it was defined by differences in either orientation or spatial frequency (with the irrelevant attribute varying randomly within a field). The results of the Walters et al. experiments were presented at ARVO but never published.

Biederman, Legge, and Luebker have embarked on an examination of how conjunctions of orientation and spatial frequency might be perceived. We have initially completed a replication of the Walters et al. experiment and are currently evaluating parametric variations. One implication of the lack of automatic texture segregation of conjunctions of orientation and spatial frequency is that the outputs of early (V1) filters are not directly perceived. Instead, they are used for subsequent processing channels. The ultimate goal of this project is to explore this linkage between early filtering and descriptors of shape that can be used for search, texture segregations, and recognition.

A Neural Net Implementation of Shape Recognition that Solves the Binding Problem Through Temporal Correlation (Hummel & Biederman)

Hummel and Biederman have developed a neural net implementation of Recognition-by-Components (RBC) which has been accepted for publication. This work is summarized in the abstract and a copy of the ms. is enclosed with this progress report.

Abstract. Upon exposure to a single view of an object, the human can readily recognize that object from any other view that preserves the parts in the original view. Experimental evidence suggests that this fundamental capacity reflects the activation of a viewpoint invariant structural description specifying the object's parts and the relations among them. This paper presents a neural network model of the process whereby a structural description is generated from a line drawing of an object and used for object classification. The model's capacity for structural description derives from its solution to the *dynamic binding problem* of neural networks: Independent units representing an object's parts (in terms of their shape attributes and interrelations) are bound *temporarily* when those attributes occur in conjunction in the systems input. Temporary conjunctions of attributes are represented by synchronized (or phase locked) oscillatory activity among the units representing those attributes. Specifically, the model uses phase locking to: a) parse images into their constituent parts; b) bind together the attributes of a part; and c) determine the relations among the parts and bind them to the parts to which they apply. Because it conjoins independent units temporarily, dynamic binding allows tremendous economy of representation, and permits the representation to reflect the attribute structure of the shapes represented. The model's recognition performance is shown to conform well to empirical findings.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review*. In press.

Developing a Front End for RBC (Biederman & Fiser)

The implementation of RBC described in the previous section takes a line drawing as an input (actually the coded end points of the lines). Irving Biederman and Jozef Fiser have embarked on developing an input system that would be able to take a grey level photograph as an input and

derive the layer two attributes (vertices, axes, and blobs) in the Hummel and Biederman implementation. The initial effort is using a mesh of Gabor "jets" developed by van der Malsberg to do an initial filtering of the image at a variety of scales and orientations.

Psychophysics of Complex Auditory Signals (Viemeister)

During this reporting period two major experimental projects were initiated. The first, with Alberto Recio, is an examination of the psychophysical properties of fractal sounds. Two types are being examined: fractional Brownian motion and fractal point processes.

Fractional Brownian motion (fBm) with parameter H ($0 < H < 1$), is defined by Mandelbrot and van Ness (SIAM Review, vol.10 No.4, 1968). It can be thought of as the output of a filter, having an impulse response equal to $t^{(H - 0.5)}$, to Gaussian white noise. The increments of fBm are stationary and self-similar with parameter H , yet fBm is not itself a stationary process. Its "spectral density" is proportional to $1/(f^{(2H+1)})$. The immediate question is whether the auditory system is sensitive to changes in the parameter H and how that sensitivity might be mediated: It is more likely that changes in the spectral shape, rather than temporal self-similarity, is the basis for discriminating changes in H . Discrimination thresholds are measured using a 3IFC task. The level of each stimulus is randomized and a high-pass noise is added to each stimulus to minimize loudness cues. It appears the the discrimination performance is predicted by a simple single channel model that simply measures the spectral density at 1k Hz. Thus, it appears that neither spectral shape nor self-similarity is used for discrimination of this class of fractal sounds. We are just beginning investigation of the discrimination of fractal point processes.

The second project, with Paul Chang, is an examination of the "temporal window" and of the processing strategies used in the detection of signals that are presented at unpredictable times. The basic experiment is a temporal analog of the probe-signal procedure developed by Greenberg and Larkin [J. Acoust. Soc. Am. 44 (1968)]. The signal is a 20 ms, 1 kHz tone and is presented in continuous broadband noise. On 80% of the trials the signal is presented synchronously with a light flash and/or a click presented to the contralateral ear. On the remaining "probe" trials the signal is presented asynchronously. As expected, detection performance is best for synchronous signals and deteriorates monotonically as the delay between the signal and marker cue increases. The temporal window measured in this way is surprisingly broad, of the order of 100's of ms. This indicates that the observer is not using the high degree of temporal resolution of which he is capable, and suggests a processing strategy in which processing effort or attention is distributed over time. We are currently measuring resolution performance in a similar task and will attempt to develop a quantitative account that relates resolution and the temporal window measured in the probe procedure.

Perceptual Processing of Reflectance Contours (Knill & Kersten)

Knill & Kersten have are extending their studies of the perceptual processing of reflectance contours, particularly in relationship to the perception of surface shape. The general hypothesis is that the visual system assumes figural regularity of reflectance edges in the world, but relative to the shapes of the surfaces on which the edges lie, what they have referred to as geodesic regularity. They are nvestigating whether such figural constraints play a role in reflectance contour coding. The general paradigm to be used is one in which efficiency measures are obtained for the discriminability of images containing smoothly curved surfaces with reflectance edges of different shapes. The images presented to subjects will be corrupted by white luminance noise, and human performance compared to that of an ideal observer to obtain a measure of statistical efficiency. The experimental hypothesis is that discrimination efficiency will be best for contours which display geodesic regularity relative to the surface in a test image. The null hypothesis is that efficiency will not vary with contour regularity, or will vary as a function of contour regularity in the image plane (independent of the shape of the surfaces from which the contours project).